

Title page

Struggling and juggling: A comparison of student assessment loads across research and teaching-intensive universities

First Author

Dr Carmen Tomas

Teaching and Learning Directorate

University of Nottingham

Nottingham

NG7 2RD

Phone: 0115 7484772

Carmen.Tomas@nottingham.ac.uk

Corresponding author

Professor Tansy Jessop

Solent Learning and Teaching Institute

Southampton Solent University

East Park Terrace

Southampton

SO14 0YN

Phone: 0238 2016982

Tansy.Jessop@solent.ac.uk

Biographical note:

Dr Carmen Tomas is Assessment Adviser in the University of Nottingham's Teaching and Learning Directorate. She has used and adapted the TESTA methodology to help improve assessment design on various programmes at Nottingham.

Professor Tansy Jessop leads 'Transforming the Experience of Students through Assessment' (TESTA) which has been used in a large number of UK universities to improve assessment and feedback across whole programmes. Tansy is Professor of Research Informed Teaching at Southampton Solent University.

Struggling and juggling: A comparison of student assessment loads across research and teaching-intensive universities

Carmen Tomas and Tansy Jessop

Learning and Teaching Directorate, University of Nottingham

Solent Learning and Teaching Institute, Southampton Solent University

Abstract

In spite of the rising tide of metrics in UK higher education, there has been scant attention paid to assessment loads, when evidence demonstrates that heavy demands lead to surface learning. Our study seeks to redress the situation by defining assessment loads and comparing them across research-and teaching intensive universities. We clarify the concept of ‘assessment load’ in response to findings about high volumes of summative assessment on modular degrees. We define assessment load across whole undergraduate degrees, according to four measures: the volume of summative assessment; volume of formative assessment; proportion of examinations to coursework; number of different varieties of assessment. All four factors contribute to the weight of an assessment load, and influence students’ approaches to learning. Our research compares programme assessment data from 73 programmes in 14 UK universities, across two institutional categories. Research-intensives have higher summative assessment loads and a greater proportion of examinations; teaching-intensives have higher varieties of assessment. Formative assessment does not differ significantly across both university groups. These findings pose particular challenges for students in different parts of the sector. Our study questions the wisdom that ‘more’ is always better, proposing that lighter assessment loads may make room for ‘slow’ and deep learning.

Key words: Assessment load; programme assessment; deep learning; summative; formative

Why assessment load is important

In spite of studies hinting at the problem of assessment load (Lizzio, Wilson and Simons 2002; Lizzio and Wilson 2013; Harland et al. 2015), the concept has not been defined in the literature. We know that all is not well with assessment and feedback. In the UK, it

manifests itself in relatively low National Student Survey scores compared to teaching quality and overall satisfaction scores. In this article, we argue that the problem is a systemic and structural one, partly the result of the segmented structure of the curriculum. Since the 1980s modular degree structures have become commonplace in the UK and many other higher education systems, with students taking concurrent credit-bearing modules each semester or term, and accumulating transferable credits towards their degrees. On the one hand, modules offer a diverse range of choice to students; on the other, they are often stand-alone offerings which do not connect well with the wider curriculum (Jessop, McNab and Gubby 2012). The modular system has contributed to an ‘assessment arms race’, characterised by frequent and fragmented summative assessment, which entices students to work in response to ‘pedagogies of control’ which elicit effort in return for grades (Harland et al. 2015; Wass et al. 2015). The result is mutual dissatisfaction, symptomatic of the “fractured and impoverished dialogue” students decry in their feedback (Nicol 2010). Academics complain of over-assessment and huge marking loads (Tuck 2012); students display high stress levels, with bunched deadlines contributing to surface approaches to learning. Academics struggle to encourage reluctant students to complete formative tasks; for students, formative tasks compete with graded ones (Jessop, El Hakim and Gibbs 2014). Lecturers bemoan the long hours spent crafting feedback which often yields peremptory glances from students more interested in their grades (Tuck 2012); and students find that feedback lands on the doorstep once the modular door has shut behind them (Price et al. 2010). Varieties of assessment come thick and fast and often in random order, confusing students with their apparently fickle requirements. These are the realities of the assessment environment many students find themselves in at UK universities (Jessop, El Hakim and Gibbs 2014; Jessop and Tomas 2017).

Research has demonstrated that high volumes of summative assessment counteract deep learning (Lizzio, Wilson and Simons 2002). A surface approach to learning is characterised by memorisation of content in isolation, disconnected from the wider body of knowledge. In contrast, deep approaches engage students in meaningful learning which integrates knowledge in the field (Marton and Saljo 1976). The modular degree context with its segmentation of the curriculum promotes a tendency to surface learning. This is compounded by high volumes of summative assessment which are characteristic of modular systems. Modular degrees contribute to a relentless diet of summative assessments, feeding grade-orientation, as lecturers plan assessments on individual modules, while students

experience them across multiple modules. Conversely there are very low ratios of formative assessment (Wu and Jessop 2018). One plausible explanation for students taking a surface approach to learning from assessment is that they focus instrumentally on achieving grades in the absence of formative assessment, attending mainly to work that ‘counts’ (Jessop, El Hakim and Gibbs 2014; Harland et al. 2015; Wass et al., 2015; Jessop and Tomas 2017); another is that heavy cognitive loads interfere with deep approaches to learning (Lizzio, Wilson and Simons 2002). Grade-orientation is a persistent problem, as demonstrated by classic studies from the 1960s and 70s in the USA and the UK (Becker, Greer and Hughes 1968; Miller and Parlett 1974). Mass higher education and marketization have amplified grade-orientation, causing students to focus more on value for money and degree outcomes than the process of learning (Arum and Roksa 2011).

Within this environment, students are having to juggle multiple assessment tasks simultaneously, leading to high cognitive load. Kalyuga (2011) categorises cognitive load as either intrinsic, linked to how students perceive the complexity of an assessment task and evaluate their capacity to complete it; or extraneous, placed on a student through factors beyond the assessment itself, for example managing workload and timing and the clarity of assessment standards. Given that cognitive loads are additive, “the focal question is not just ‘what type of assessment’ but ‘what type of assessment system’ students are experiencing on both cognitive and affective levels” (Lizzio and Wilson 2013, 391). The concept of assessment load has a direct bearing on students’ cognitive load, particularly given this additive dimension.

Until recently, most assessment and feedback interventions have focused on lecturers’ practice on individual modules, or tackled one feature of the complex system of which assessment is an interdependent part (Gibbs and Dunbar-Goddet 2009; Gibbs 2013). Inevitably, the results of this focus on single aspects have been disappointing, because “within complex systems...the probability is that any change may wobble the dynamics but no more” (Knight 2002, 112). There has been a shift in the focus from modular attempts to change assessment practice, to more strategic and holistic approaches focussing on programme-level assessment. These take a more systemic approach to change, often articulating with quality assurance frameworks and addressing the structure of degrees (Bloxham and Boyd 2007; Gibbs and Dunbar-Goddet 2009; PASS Project (2009-12); Jessop, McNab and Gubby 2012; Jessop, El Hakim and Gibbs 2014).

Universities using the ‘Transforming the Experience of Students through Assessment’ (TESTA) method have collected in-depth data about programme assessment environments since 2009 on degree programmes in more than 50 universities, mainly in the UK, but also in Australia and India. We have drawn on TESTA evidence to define four features of assessment which comprise assessment load in the eyes of students and lecturers. These are: first, the volume of summative assessment which students experience over three years of a degree. Summative assessment is designed mainly to measure student achievement through grading, but may have learning benefits for students (Carless 2007). It is the most obvious and clearest measure of assessment load. Second, the volume of formative assessment, designed to fine tune students’ understanding of concepts and standards, which is often less clearly understood (Sadler 1989; Black and Wiliam 1998; Boud 2000). TESTA defines formative assessment as a requirement which prompts feedback and is ungraded. Both formative and summative assessment add to students’ assessment load through the time and cognitive load involved in completing tasks, and in the case of summative, performance anxiety about assessment, and emotional responses to feedback (Boud and Falchikov 2006; Pitt and Norton 2017). Third, the proportion of examinations, as these are usually high stakes, often testing and integrating a wider domain of knowledge than coursework, requiring extensive preparation, and adding to the challenge and level of demand. Fourth, the number of different varieties of assessment to which students are exposed on a degree programme. These require students to master unfamiliar, often creative and randomly sequenced modes of assessment. They contribute to assessment load in subtle ways, for example making it more difficult for students to gauge the standards expected or to use previous feedback.

Our study compares assessment loads across two broad university types: research-intensive and teaching-intensive. Research intensive universities evidence a strong research culture through impacts, outputs and the research environment, often sustained by external or government funding linked to research performance. Research-rich universities prioritise resources and time for research, with academic staff generally having lighter teaching loads and more research time than new ‘teaching-intensive’ universities (Fung, Besters-Dilger, van der Baart 2017). Entry tariffs for students are high, and students require less academic support to reach the standards (Gibbs and Dunbar-Goddet 2009). By teaching-intensive, we mean universities where teaching is the core business and the research environment is less prominent within the culture and activities of the institution. Typically, in teaching-intensive

universities, academics teach for 18 hours a week during term-time; the intake of students is more diverse and entry tariffs for students are lower (McLean, Abbas and Ashwin 2013). Consequently, more academic time is spent supporting students from diverse educational backgrounds to reach academic standards, than on undertaking research (Ainley 2008).

Literature on assessment load has mainly addressed generic sector trends (Jessop and Tomas 2017). Investigations of institution types are scarce. Early studies drew some comparisons based on very small samples (Gibbs and Dunbar-Goddet 2007; 2009). More recently, the Student Academic Experience Survey (SAES) has also introduced a comparison of aspects of assessment across different university types (Neves and Hillman 2016; 2017). While the SAES data is based on large-scale survey responses from circa 15,000 students across the UK, TESTA data provides a richer and more robust picture of assessment on degree programmes. The concept of assessment load discussed here is fuller, making room for a more textured account of students' lived experience than SAES' narrower focus on summative and formative loads. Methodologically, the data from TESTA is more robust, when compared to student self-reported data in the SAES, as it has been collected through a clarifying conversation, recorded in the TESTA audit, and triangulated with two other methods (the Assessment Experience Questionnaire and student focus groups). Additionally TESTA evidence has a strong focus on understanding whole programme data, whereas the SAES asks students to calculate how much summative and formative assessment they have experienced per 'term' which may not always be self-explanatory or comparable across different lengths of terms and semesters. In order to characterise assessment loads in different institution types, evidence needs to consist of more than the fallible recollections of students about the number of formative and summative tasks per term.

Ethics

In each of the 14 universities represented in this study, the TESTA methodology has undergone scrutiny to ensure that the research conformed to ethical protocols. Ethical scrutiny ensures that research participants have the right to withdraw, that the data is confidential, and that participants remain anonymous. The developmental purpose of TESTA frames the research process, so that programme teams make decisions based on evidence without any managerial influence or interference. The analysis of TESTA data honours the

contract with research participants since all universities and programmes have been anonymised.

Measuring assessment load: overall and across different university types

In this section, we outline an approach for measuring assessment load which is based on the audit element of the TESTA methodology (Jessop and Tomas 2017). The audit consists of a dialogic interview with the programme leader over the programme documents which constitute the planned curriculum. The audit elicits programme-level data about the balance of formative and summative assessment, varieties of assessment, and the proportion of assessment which takes place by examination. The examination measure calculates the proportion of the total summative assessment tasks which take place by means of an examination, rather than the weighting of marks from examinations.

In general, the programme documents specify the number and types of summative assessment on each module, but are usually more opaque about ungraded formative assessment tasks. However, the qualitative nature of the audit enables it to go beyond the planned curriculum in order to clarify instances of formative assessment in conversation with the programme leader. The uniqueness of the audit is that it captures assessment data across whole programmes rather than modules only. Overall, it collects data about:

- Volume of summative assessments
- Volume of formative assessments
- Number of varieties of assessment
- Proportion of tasks by examination
- Amount of written feedback (through sampling post-audit)

The original audit methodology measured features of assessment and explored the implications of the balance and load of assessment in different university types on a small-scale, comparing programmes in one new university, one red-brick research-intensive, and Oxbridge (Gibbs and Dunbar-Goddet 2007; 2009). This earlier study paved the way for a larger scale study which characterised features of assessment based on statistical methods using TESTA data (Jessop and Tomas 2017). In the current study, we use these categories of low, medium and high to compare assessment loads across research intensive and teaching

intensive universities, with an equal split across 14 universities. Of the 73 TESTA programmes represented, 35 are within research-intensive and 38 are in teaching-intensives.

Data analysis

The data analysis has two aspects: first, we describe overall findings across both university groups using interquartile ranges; second, we compare TESTA data from research-intensive and teaching-intensive universities using statistical tests of difference. Given the non-parametric nature of the TESTA data, we have decided to use the more resilient measure of interquartile ranges and medians to characterise features of assessment accurately rather than using the parametric equivalent of means. The analysis uses basic descriptive measures to offer insights into the range, dispersion, and central tendency. All quartile boundaries are classified in terms of High, Medium and Low boundaries:

- Low: below first quartile boundary
- Medium: above first quartile and below third quartile. Medians are indicated for further information on the second quartile.
- High: above third quartile boundary

In addition to presenting the descriptive analysis, we have used Mann-Whitney U tests to identify significant differences in assessment load across the two university groups. The Mann-Whitney test is a non-parametric test for independent samples. We have used the two-tailed modality since there is no theoretical rationale to expect a particular direction in which differences might occur.

Findings across both university types

The findings build on previously published analysis about assessment patterns using the whole data set discussed here (Jessop and Tomas 2017). Table 1 outlines these patterns using interquartile ranges.

Table 1. High, medium & low IQRs (n = 73 programmes) (Jessop & Tomas 2017)

Characteristic	Low	Medium	High
Volume of summative assessment	Below 33	33-48	More than 48
Volume of formative only	Below 1	1-19	More than 19
% of marks from examinations	Below 10%	10-30%	More than 30%
Variety of assessment methods	Below 8	8-15	More than 15

Comparing assessment loads across research and teaching-intensive universities

In this study we use TESTA audit results to compare medium ranges and median data from seven research-intensive universities (n=35 programmes) and seven teaching-intensive universities (n=38 programmes). Table 2 compares ‘medium’ assessment loads in research and teaching intensive universities According to Mann-Whitney test results: an asterisk (*) indicates where differences are significant; where the null hypothesis is retained this is indicated as n.s. (not significant). Statistical tests of difference using Mann-Whitney U tests (two-tailed) indicate the following:

- Teaching-intensive institutions have a lower summative assessment load (median = 35) than the research-intensive group (median = 50). The distributions of these two groups differed significantly (Mann-Whitney U, $Z = 4.367$, $p < 0.00001$, two tailed).
- Formative assessment load in research (Median=3) and teaching intensive (Median=7) is not significantly different (Mann-Whitney U, $Z = 0.96066$, $p > 0.05$, two tailed).

- The proportion of assessment by examinations is significantly different and distributed differently in research-intensive universities (Median= 30%) than teaching intensive (Median = 10%) (Mann-Whitney U, $Z=-5.1235$, $p<0.00001$, two tailed).
- Variety of assessments in teaching intensive (Median = 15) and research intensive (Median=8) and the distributions in these groups differ significantly (Mann-Whitney U, $Z= 4.74255$, $p<0.00001$, two tailed).

Table 2: Comparison of medium loads in research and teaching intensive

Characteristic	Medium (Research intensive)	Medium (Teaching intensive)	Overlap RI and TI	Mann- Whitney U test results
Volume of summative assessment	41-79 [Median 50]	34-41 [Median 35]	none	Research intensive*
Volume of formative only	1-26 [Median 3]	3-17 [Median 7]	3-17	n.s.
% of marks from examinations	27-42% [Median 30%]	5-19% [Median 10%]	none	Research intensive*
Variety of assessment methods	8-10 [Median 8]	12-14 [Median 15]	none	Teaching intensive*

Discussion of assessment load findings

Formative assessment

The absence of significant differences in formative assessment loads across research and teaching-intensive universities is entirely plausible within the wider context of TESTA's

large-scale multi-method research process. Repeatedly, across both parts of the higher education sector, TESTA data evidences that academics struggle to construct meaningful formative tasks which align with the learning outcomes and summative assessments, while engaging students in more than simply ‘dashing it off’ to satisfy minimal requirements. Summative assessment competes for student time and attention, especially in modular systems with high assessment loads. Without adjusting high summative assessment loads, students find the demands of juggling both summative and formative tasks across multiple modules difficult (Jessop, El-Hakim and Gibbs 2014; Harland et al 2015; Jessop and Tomas 2017). Previous studies based on TESTA data demonstrate the challenge of designing meaningful formative across both types of university (Wu and Jessop 2018). Given that the digital sphere offers opportunities for authentic learning opportunities, many academics are responding to the challenges of designing formative assessments using real world tools such as blogging (Lombardi 2007; Barlow and Jessop 2016). These have the virtue of requiring students to read and write more productively, while not necessarily increasing assessment loads for academics.

Lack of a statistically significant difference in formative assessment loads across research-intensive and teaching-intensive universities contradicts findings in the latest two Higher Education Policy Institute (HEPI)-Higher Education Academy (HEA) Student Academic Experience Surveys (SAES) (Neves and Hillman 2016; 2017). The SAES uses self-reported data from 15,022 and 14,057 students to suggest that, on average, students at research-intensive universities experience almost double the formative load than their peers in post-1992 universities (Neves and Hillman 2016, p. 24; 2017, p.41). Based on a calculation of SAES means, which collapses Russell Group and pre-92 universities into one research-intensive category, and multiplies it to equivalence three years of a degree, students in research-intensive universities reported 19 formative assessments compared to students in teaching-intensive universities who reported 10.2 formative assessments over three years.

SAES’s finding that research-intensive universities are ahead of the game on formative assessment should be treated with caution and interpreted against the more troubling realities that TESTA data evidences across the piece, including that there is far less formative assessment than the SAES report chronicles. TESTA medians for research-intensive universities (n=3) and for teaching-intensive universities (n=7) paint a starker picture of students’ experience of formative assessment.

Significant differences in summative assessment

Our analysis indicates that research intensives (median=50) have significantly more summative assessment than teaching-intensives (median= 35). This reflects the nature of the two different types of institution, with teaching-intensives having many more applied, creative, ‘real-world’ assessment tasks which are often project or portfolio-based (Adams and McNab 2013). Contrastingly, research-intensives tend to have a more traditional assessment diet within their subject disciplines; typically, essays, examinations, laboratory reports and, in final year, research dissertations (Gibbs and Dunbar-Goddet 2009). Variations in the number of assessment tasks raises questions about the relative complexity of tasks. So for example, it is likely that TESTA’s calculation of 227 summative tasks on a Mathematics programme at a research intensive university indicates relatively brief tasks. Contrastingly, the 12 summative assessment points over a three year creative degree in a teaching-intensive university are more likely to be large tasks involving collaborative project work.

Harland et al. (2015) describe small and frequent summative assessment tasks as a ‘pedagogy of control’, catalysing student effort but not necessarily deep learning. The consequences of many summative assessment tasks colliding at similar points in the semester are (a) student non-attendance of teaching events; (b) squeezing out of formative opportunities; and (c) surface learning (Lizzio, Wilson and Simons 2002; Wu and Jessop 2018). More summative assessment is not necessarily better for student learning because students are working harder, as effort does not necessarily equate to depth (Harland et al. 2015). Students may be working more frequently, but in more instrumental and superficial ways. Contrastingly, it is not necessarily true that large scale collaborative project work in teaching-intensive always engages students in deep learning. Hanney and Savin-Baden (2013) problematise project-based learning as linear, directive and closed. These qualities of project-based tasks may leads to students taking a surface approach to learning rather than a deep approach.

Significant differences in volume of examinations

Students in research-intensive universities experience significantly more assessment by examinations (median 30%) than their counterparts in the teaching-intensive part of the

sector (median 10%). This may be related to more traditional subjects and disciplines represented in research-intensive universities, which are well-suited to written forms of assessment, than the creative, professionally-oriented, and applied subjects, characteristic of teaching-intensive universities. There is a higher proportion of STEM subjects in research-intensives, which use more examinations in their assessment regimes (Warren Piper et al 1996; Pryor and Crossouard 2010; Jessop and Maleckar 2016). The student demography in research and teaching-intensive universities differs too, with more students with declared disabilities in teaching-intensive universities, and of those, a large proportion with Specific Learning Differences (Higher Education Funding Council for England 2015). Examinations are one of the least inclusive modes of assessment, providing an explanation for their lower emphasis in teaching-intensive universities. A further factor may be that teaching-intensive universities have lower entry tariffs for students. Examinations are a high stakes form of assessment, with students achieving lower marks on examinations than by coursework (Bridges et al. 2002; Richardson 2015). Given the risk of failure among students whose educational achievements are generally lower on entry than their research-intensive student counterparts, teaching-intensive universities may deliberately avoid them. Finally, examinations are relatively plagiarism proof, which may go some way to explaining their use in disciplines dominated by essays and written modes of assessment, the mainstay of research-intensive universities. Plagiarism is less of an issue with project-based learning and in real-world, applied and professional learning contexts, where students are interacting with, and judged both by academics and industry professionals, as part of the assessment process.

Traditionally, examinations have been regarded as vehicles for memorisation and regurgitation of isolated facts, characteristic of surface approaches to learning (Entwistle and Entwistle 1991; Newstead and Findlay 1997). Yet, well-designed examinations which allow students time to prepare and to integrate their learning before sitting them, may encourage deep learning (Gibbs and Simpson 2004). At the same time, students prefer coursework assessment, which tends to produce more meaningful outcomes than examinations, particularly when the assessment is outward facing and has a real audience (Fung 2017). While traditionalists may argue that examinations are a hallmark of more rigorous degree qualifications, the evidence is equivocal, particularly in an environment of student participation, co-creation and research-based learning (Brew 2010; Bovill et al. 2011).

Significant differences in varieties

The variety of assessments is significantly higher in teaching intensive universities (median 15) than in research-intensive universities (median 8). Explanations for higher varieties of assessment relate to the disciplinary fold of teaching-intensive universities, where diverse and applied fields of study have learning outcomes not easily or solely able to be demonstrated in traditional written formats. Expansive varieties of assessment such as posters, podcasts, films, design projects, feature articles, film scripts, reflective journals, and vivas reflect the wide-ranging scope of applied courses. Teaching-intensive universities recruit higher proportions of students with specific learning differences such as dyslexia, driving a range of more inclusive modes of assessment.

In theory, high varieties of assessment have the virtue of being more inclusive, potentially enabling students with learning differences to thrive. In practice, TESTA data shows that high varieties of assessment are not an unalloyed virtue, because when, as often happens, they are randomly sequenced through modular degrees, they tend to confuse students, preventing them from internalising goals and standards (Jessop and Tomas 2017). In the humanities, Oxbridge students write at least an essay every week of term, and write essays in their exams. It is not surprising that they learn how to write essays really well. Contrastingly, students who “bump into all kinds of curious assessments demands” find it much harder to express their learning effectively (Gibbs 2017, 64). They are often “bewildered by the variety and idiosyncrasy of assignment demands” (Ibid., 58). A balance between inclusion, creativity and a certain amount of repetition of assessment types needs to be struck in assessment design.

Conclusion

Our study has defined and characterised student assessment load in undergraduate study programmes by university type and compared loads across research and teaching intensive groups. It adds insight to, and challenges existing research on the balance between formative and summative tasks across different university types (Neves and Hillman 2016; 2017). Using TESTA audit data we have shown that there are no significant differences between research and teaching-intensive universities in formative assessment loads. Research intensive universities have higher summative assessment loads and a higher proportion of examinations in their assessment patterns, when compared with the teaching intensive group.

Teaching intensive institutions have a broader range of varieties of assessment than research intensive universities.

Statistical differences between research and teaching-intensive universities in measures of summative assessment, examinations and varieties of assessment are not surprising, given the types of courses and different emphases across the two university types. Highly competitive research-intensive universities with a high proportion of convergent disciplines in STEM fields, have shaped assessment environments around high summative assessment loads and examinations. Teaching-intensive universities, with their varied, creative and applied courses offer a wider variety of assessment types to students. In the discussion we posed questions about the virtue of more summative, higher proportions of examinations and wider varieties of assessment, arguing that ‘more’ may prompt less deep learning, contributing to surface approaches through high summative and examination loads, and sowing confusion through randomly sequenced varieties of assessment.

The surprising conclusion from our study is that there are no statistical differences in formative assessment loads across the two university groups. Across both university types, within their modular systems, students are struggling with and juggling assessment loads. More summative assessment, more examinations, and more varieties all run the risk of feeding instrumental and surface approaches to learning, where students are forced to deploy tactical strategies to cope with complex assessment loads. Paradoxically, the solution appears to lie in a culture shift which embraces less summative assessment, fewer examinations and pared down varieties across whole programmes, in favour of more creative, meaningful and engaging formative tasks which help students to learn. Assessment design needs to attend to the resurgence of the slow learning movement, encouraging students to engage in formative tasks for pleasure, for learning, and out of curiosity borne of scholarship and inquiry (Berg and Seeber 2016). The only way to take these radical steps is together, through a whole programme approach, and a climb down from the ‘assessment arms race’ (Harland et al. 2015).

Disclosure statement

No potential conflict of interest was reported by the authors.

References

- Adams, J., and N. McNab. 2013. Understanding Arts and Humanities Students' Experiences of Assessment and Feedback. *Arts and Humanities in Higher Education* 12(1): 36-52. DOI: 10.1177/1474022212460743
- Ainley, P. 2008. "The Varieties of Student Experience – An Open Research Question And Some Ways To Answer It." *Studies in Higher Education* 33(5): 615-624. DOI: 10.1080/03075070802373222
- Arum, R., and J. Roksa. 2011. *Academically Adrift: Limited Learning on College Campuses*. Chicago, IL: University of Chicago Press.
- Barlow, A. and T. Jessop. 2016. "You can't write a load of rubbish": Why blogging works as formative assessment. *Educational Developments*. 17(3), 12-15. SEDA.
- Becker, H.S., Greer, B. and Hughes, E.C. 1968. *Making the grade: The academic side of college life*. New York: Wiley.
- Berg, M. and Seeber, B. 2016. *The Slow Professor: Challenging the Culture of Speed in the Academy*. Toronto. University of Toronto Press.
- Black, P., and D. Wiliam. 1998. *Inside the Black Box: Raising Standards through Classroom Assessment*. London: Grenada Learning.
- Bloxham, S., and P. Boyd. 2007. *Developing Effective Assessment in Higher Education: A Practical Guide*. Maidenhead: Open University Press.
- Boud, D. 2000. "Sustainable Assessment: Rethinking Assessment for the Learning Society." *Studies in Continuing Education* 22 (2): 151–167. 10.1080/713695728
- Boud, D., and N. Falchikov. 2006. "Aligning Assessment with Long-term Learning." *Assessment & Evaluation in Higher Education* 31 (4): 399–413. doi:10.1080/02602930600679050.
- Bovill, C., Cook-Sather, A., and Felten, P. 2011. "Students as co-creators of teaching approaches, course design and curricula: implications for academic developers". *International Journal for Academic Development*, 16 (2). pp. 133-145. ISSN 1360-144X
- Brew, A. 2010. "Imperatives and challenges in integrating teaching and research". *Higher Education Research and Development*. 29:2, 139-150, DOI:10.1080/07294360903552451
- Bridges, P., A. Cooper, P. Evanson,, C. Haines, D. Jenkins, D Scurry, H. Woolf, and M. Yorke. 2002. "Coursework marks high, examination marks low: discuss." *Assessment and Evaluation in Higher Education* 27(1): 35-48
- Carless, D. 2007. "Learning-oriented Assessment: Conceptual Bases and Practical Implications." *Innovations in Education and Teaching International* 44 (1): 57–66. doi:10.1080/14703290601081332.
- Entwistle, N.J. & Entwistle, A. 1991. Contrasting forms of understanding for degree examinations: the student experience and its implications. *Higher Education* 22: 205. DOI:10.1007/BF00132288
- Fung, D., J. Besters-Dilger, and R. van der Vaart. 2017. Excellent Education in Research-Rich Universities. Leuven: League of European Universities (LERU) Position Paper.
- Fung, D. 2017. *A Connected Curriculum for Higher Education*. London. UCL Press.
- Gibbs, G. 2017. Alternative explanations for high and low average marks. *Capture 5: Spring 2017*, 55-67. At: https://issuu.com/theuniversityofwinchester/docs/capture_20vol._205?e=2137537/55140691
- Gibbs, G. 2013. Reflections on the changing nature of educational development. *International Journal for Academic Development*. 18:1, 4-14, DOI: 10.1080/1360144X.2013.751691

- Gibbs, G. and Dunbar-Goddet, H. 2007. The effects of programme assessment environments on student learning, York: Higher Education Academy.
- Gibbs, G., and H. Dunbar-Goddet. 2009. "Characterising Programme-level Assessment Environments that Support Learning." *Assessment & Evaluation in Higher Education* 34 (4): 481–489. doi:10.1080/02602930802071114
- Gibbs, G., and C. Simpson. 2004. "Condition under Which Assessment Supports Students' Learning." *Learning and Teaching in Higher Education* 1: 3–31.
- Hanney, R. and M. Savin-Baden. 2013. The problem of projects: understanding the theoretical underpinnings of project-led PBL. *London Review of Education*. Vol. 11, No. 1, 7–19, <http://dx.doi.org/10.1080/14748460.2012.761816>
- Harland, T., A. McLean, R. Wass, E. Miller, and K. N. Sim. 2015. "An Assessment Arms Race and Its Fallout: High-stakes Grading and the Case for Slow Scholarship." *Assessment & Evaluation in Higher Education*. 40 (4): 528–541. doi:10.1080/02602938.2014.931927.
- HEFCE. 2015. Support For Higher Education Students With Specific Learning Difficulties. Available at: <http://www.hefce.ac.uk/pubs/rereports/year/2015/spld/>
- Jessop, T., Y. El Hakim, and G. Gibbs. 2014. "The Whole is Greater Than the Sum of Its Parts: A Large-scale Study of Students' Learning in Response to Different Programme Assessment Patterns." *Assessment & Evaluation in Higher Education* 39 (1): 73–88. doi:10.1080/02602938.2013.792108.
- Jessop, T., and B. Maleckar. 2016. "The Influence of Disciplinary Assessment Patterns on Student Learning: A Comparative Study." *Studies in Higher Education* 41 (4): 696–711. doi:10.1080/03075079.2014.943170.
- Jessop, T., N. McNab, and L. Gubby. 2012. "Mind the Gap: An Analysis of How Quality Assurance Procedures Influence Programme Assessment Patterns." *Active Learning in Higher Education* 13 (2): 143–154. doi:10.1177/1469787412441285.
- Jessop, T., and C. Tomas. 2017. "The Implications of Programme Assessment Patterns for Student Learning." *Assessment & Evaluation in Higher Education* 42 (6): 990–999. doi:10.1080/02602938.2016.1217501.
- Kalyuga, S. 2011. "Cognitive Load Theory: How Many Types of Load Does It Really Need?" *Educational Psychology Review* 23 (1): 1–19. doi:10.1007/s10648-010-9150-7.
- Knight, P. T. 2002. "Summative Assessment in Higher Education: Practices in Disarray." *Studies in Higher Education* 27 (3): 275–86. doi:10.1080/03075070220000662
- Lizzio, A., Wilson, K. and Simons, R. 2002. University students' perceptions of the learning environment and academic outcomes: Implications for theory and practice. *Studies in Higher Education*. 27: 27–51.
- Lizzio, A., and K. Wilson. 2013. "First-year Students' Appraisal of Assessment Tasks: Implications for Efficacy, Engagement and Performance." *Assessment & Evaluation in Higher Education* 38 (4): 389–406. doi:10.1080/02602938.2011.637156.
- Lombardi, M. 2007. Authentic Learning for the 21st Century: An overview. Educause. At: <https://net.educause.edu/ir/library/pdf/ELI3009.pdf>.
- Marton, F., and R. Saljo. 1976. "On Qualitative Differences in Learning I: Outcome and Process." *British Journal of Educational Psychology* 46: 4–11.
- McLean, M., A. Abbas, & P. Ashwin. 2013. "The Use and Value of Bernstein's Work in Studying (in) Equalities in Undergraduate Social Science Education." *British Journal of Sociology of Education* 34 (2): 262–280. doi:10.1080/01425692.2012.710007
- Miller, C. M. L., and M. Parlett. 1974. *Up to the Mark: A Study of the Examination Game*. London: Society for Research into Higher Education.

- Neves, J., and N. Hillman. 2016. The 2016 Student Academic Experience Survey. Oxford: The Higher Education Policy Institute (HEPI) and The Higher Education Academy (HEA).<http://www.hepi.ac.uk/2016/06/09/hepi-hea-2016-student-academic-experience-survey/>.
- Neves, J., and N. Hillman. 2017. 2017 Student Academic Experience Survey. York: Higher Education Academy.
- Newstead, S. E. and Findlay, K. 1997. "Some problems with using examination performance as a measure of teaching ability". *Psychology Teaching Review*, 6, 14-21.
- Nicol, D. J. 2010. "From Monologue to Dialogue: Improving Written Feedback Processes in Mass Higher Education." *Assessment & Evaluation in Higher Education* 35 (5): 501–517.10.1080/02602931003786559
- PASS Project 2009-2012. Programme Assessment Strategies. At: <http://www.pass.brad.ac.uk>
- Pitt, E. and Norton, L. 2017. 'Now that's the feedback I want!' Students' reactions to feedback on graded work and what they do with it. *Assessment & Evaluation in Higher Education* 42 (4): 499-516.
- Price, M., K. Handley, J. Millar, and B. O'Donovan. 2010. "Feedback: All That Effort, but What is the Effect?" *Assessment & Evaluation in Higher Education* 35 (3): 277–289. doi:10.1080/02602930903541007.
- Pryor, J., and B. Crossouard. 2010. "Challenging Formative Assessment: Disciplinary Spaces and Identities." *Assessment & Evaluation in Higher Education* 35 (3): 265–76. doi: 10.1080/02602930903512891
- Richardson, J. 2015. "Coursework Versus Examinations In End-Of-Module Assessment: A Literature Review". *Assessment & Evaluation in Higher Education* 40 (3): 439-455.
- Sadler, D. R. 1989. "Formative Assessment and the Design of Instructional System." *Instructional Science* 18: 119–144.10.1007/BF00117714
- Tuck, J. 2012. "Feedback-Giving As Social Practice: Teachers' Perspectives an Feedback as Institutional Requirement, Work and Dialogue." *Teaching in Higher Education* 17(2): 209-221.
- Warren Piper, D., D. D. Nulty, and G. O'Grady. 1996. Examination Practices and Procedures in Australian Universities. Higher Education Division, Department of Employment Education and Training. Evaluations and Investigations Program. AGPS, Canberra.
- Wass, R., T. Harland, A. McLean, E. Miller, and K. N. Sim. 2015. "Will Press Lever for Food: Behavioural Conditioning of Students through Frequent High-Stakes Assessment." *Higher Education Research and Development* 34 (6): 1324–1326.10.1080/07294360.2015.1052351
- Wu, Q. and Jessop, T. 2018. "Formative assessment: missing in action in both research-intensive and teaching focused universities?" *Assessment & Evaluation in Higher Education*. Published online 15 January 2018.

